INTRODUCTION

Agriculture is crucial for India's economy. Predicting crop yields helps farmers and policymakers make better decisions. Our project will use data science techniques to forecast crop yields in India using the "Crop Yield Data India" dataset from 1997.

DATASET OVERVIEW

The dataset includes:

- Types of crops and their production levels
- Area under cultivation
- Season of cultivation
- State-specific information
- Annual rainfall, fertilizer use, pesticide use, and yield

PROBLEM STATEMENT

The primary objective is to develop a predictive model that accurately forecasts crop yields based on relevant factors such as rainfall, fertilizer application, and pesticide usage. This will help identify key factors affecting crop production.

METHODOLOGY

- **Data Preprocessing:**
  Clean the data using Pandas.
  Handle missing values and outliers.
- **Exploratory Data Analysis (EDA):**
  Use Matplotlib, Pandas, Seaborn, and Plotly for data visualization.
  Identify trends and patterns.
- **Feature Engineering:**
  Select important features for prediction.
  Create new features if needed.
- **Model Development:**
  Split data into training and testing sets.
  Use Scikit-Learn to develop models like Linear Regression.
  Adjust the models for better accuracy.
- **Model Evaluation:**
  Evaluate models using metrics like R-squared and Mean Squared Error.
- **Interpretation:**
  Analyze results to provide insights.

EXPECTED OUTCOMES

- A model that accurately predicts crop yields.
- Insights into factors affecting crop yields.
- Recommendations for better farming practices.

CONCLUSION

By predicting crop yields, our project will help improve agricultural practices and support decision-making in agriculture. We will apply the skills learned in our course, such as data cleaning, visualization, and machine learning, to achieve our goals.